数理解析研究所講究録
1041巻 1998年 35-38

確率的分散アルゴリズムに対する ネットワークのサイズに関する情報について

坂本 直志 (Naoshi SAKAMOTO)
東京工業大学 大学院情報理工学研究科
<sakamoto@noc.titech.ac.jp>

要旨

匿名ネットワーク上の確率的分散アルゴリズムにおいて、誤りを許さないモデルと誤りを許すモデルを考えると、各頂点に異なる番号を与えるために必要な情報は異なる。誤りを許さないモデルでは、「頂点数」という情報では異なる番号を振ることができるが、「頂点数の上限」という情報では異なる番号を振ることができない。

本研究では、自然数の部分集合を与え、その要素のうちのどれかと頂点数の積を与えるという、その集合から定義される情報を考えた。これらの情報に対して、他の集合で定義される情報へ分散アルゴリズムで誤り無しで変換可能かどうかを調べた。その結果、二つの自然数の部分集合が変換可能であるための必要十分条件を見付けた。また、「頂点数」と「頂点数の上限」との間に一方的に変換可能な情報が無限個存在することが分かった。

1 はじめに

分散アルゴリズムとは複数のコンピュータがネットワークを介して、協調して動作し、計算を行うモデルである。

一般の分散アルゴリズムでは、各コンピュータに唯一の番号が割り当てられており、その番号を使用して相互のコンピュータを区別し、通信を行ったりするが、そのような番号が与えられていないことを前提とする「匿名ネットワーク」というネットワークに関する議論もある。この匿名ネットワークに関する研究のうち、もっとも大きなテーマは、どのような条件を与えると、匿名ネットワークは自立的に各頂点に唯一の番号を振り、一般的な分散アルゴリズムの前提条件を満たすようになるかである。

匿名ネットワーク上のコンピュータに対して確率的な動作を許す時、頂点数に関する情報を与えると、各頂点に番号を振ることができることがBar-Ilan と Zernik [BZ89] により示されている。

この確率的な動作を更に分析すると確率 1 で停止し、誤りを出力しないモデルでは、全頂点に異なる番号を振るには正確に全頂点数を知る必要があるが、必ず停止し、定数 r 以下の確率で誤った結果を出力するモデルでは、頂点数の上限を知れば良いことが分かる。

そこで、この、正確な全頂点数と、頂点数の上限という情報の差を比較するため、頂点数に関するさまざまな情報を誤りを許さないモデルに与え、どのようなことが計算できるかを考える。

頂点数を N とした場合、頂点数に関する情報として関数 f を用い、$f(N)$ を匿名ネットワークに与えた場合、f^{-1} が確率的なモデルで誤り 0 で計算可能であれば、N を与えることと等価になる。例えば頂点数の5倍の値を与えることを考えると、与えられた値の1/5を計算することで、正確な頂点数が分かるので、これは正確な頂点数を与えることと、等価な情報になる。そこで、入力 N から計算される $f(N)$ の値に対して、$f^{-1}(f(N))$ の計算が一意でないような情

報として、 $f(N)$ が複数の値のうちのどれかを値として取ることを考える。例えば正確な頂点数か頂点数を2倍した数かどちらかを与えると言うような情報の場合、与えられた数が頂点数なのか、それとも頂点数の2倍なのかは数を見ただけでは決定できない。

[坂本 97] において、定数個の頂点に特別な値を与えると、その与えた定数の約数個の生成木にネットワークを分割するアルゴリズムを与えた。これを応用することにより、確率的分散アルゴリズムにおいて、頂点数を超える値が与えられる場合には、必ずネットワーク上に根付の同型な生成木を作ることができることが分かる。その結果、頂点数の定数倍を情報として与える際、生成木を作ることにより頂点数の約数は得られるので、自然数でない数の積が与えられれば、正確な頂点数を求めることが可能になる場合もある。

本研究では自然数からなる帰納的可算集合を考え、その要素のどれかと頂点数との積を情報として与えた。これらの集合に対する情報が与えられた際、別の情報となりうる値が、誤り無しで出力できる確率的な分散アルゴリズムが存在するかという条件により、情報同士の関係を調べた。その結果、そのような分散アルゴリズムが存在することの必要十分条件となる集合の性質を見付けた。

2 準備

分散アルゴリズムとはネットワーク上の頂点に同一のプログラムを与え、各頂点同士は辺に沿って互いにメッセージをやりとりし、計算するモデルである。

ネットワークをポートナンバーがつけられたグラフ $G = (V, E, \sigma)$ であらわす。但し、 V は有限集合、 $E \subseteq V \times V$, $\sigma[v]$ は $v \in V$ に接続した頂点に対して 1 から $\deg(v)$ までの値をユニークに割り当てる関数とする。

ここで、グラフの頂点は V の要素により定義されるが、実際のネットワーク上の計算ではその要素の名前など、頂点のラベルの情報は一切与えない。このように各頂点に頂点のラベルの情報を与えないネットワークは匿名ネットワークと呼ばれる。一般に、匿名ネットワーク上では各頂点が異なる出力を出すことは非常に難しい。

分散アルゴリズムはグラフの頂点上に配置され、それぞれ非同期に動く。通信のコストは一定ではなく、通信をするまで決定できない。但し、本論文では FIFO と呼ばれる、先に送ったメッセージは後から送ったメッセージより必ず先に届くと言う仮定をする。

匿名ネットワーク上の各コンピュータに与える情報は自然数とする。それらの自然数はあらかじめある条件を満たすように与え、また、その条件をあらかじめ各コンピュータは知っているものとする。そのような条件を初期条件という。初期条件はネットワークの構造から決められるもので、ネットワークのグラフにより各頂点への自然数の割当を決める関数とみなすこともできる。また、グラフの頂点に対して自然数を割り当てたようなラベルつきグラフの集合と考えることもできるが、本論文ではそのように考えた集合が帰納的可算集合であるものだけを扱う。このような初期条件を以下 recursive coloring (r.c. と略す) と呼ぶことにする。

r.c. に属し、一般的な匿名ネットワーク上でよく用いられる、初期条件を次のように定義する。

Definition 2.1
- $\mathcal{UNIQUENUMBER}$: グラフの G に対して $a \in \mathcal{UNIQUENUMBER}(G)$ はグラフの頂点に対して、各頂点に唯一の番号を与える。
- $\mathcal{SIZE}$: グラフ G に対して $a \in \mathcal{SIZE}(G)$ は全ての頂点にグラフの頂点数を与える。
- $\mathcal{UPPERBOUND}$: グラフ G に対して $a \in \mathcal{UPPERBOUND}(G)$ は各頂点にグラフの頂点数以上の有限な値を与える。

さて、確率的な状態遷移をする分散アルゴリズムのモデルとして、

- 全てのプロセッサが確率 1 で停止し、出力に誤りが無い。
- 全てのプロセッサは必ず有限時間内で停止し、全頂点の出力が条件にあわない確率が $r > 0$ 以下。

を考える。

ここで、任意のネットワークに対して、初期条件 B を満たす初期値を与えると、初期条件 A を満たす初期値を出力する計算が、全てのプロセッサが確率 1 で停止し、出力に誤りが無い分散アルゴリズムでできる時、$A \leq_{\mathrm{zp}} B$ と書くことにし、確率 $1-r$ で出力できる有限時間内で必ず停止する分散アルゴリズムでできる時、$A \leq_{\mathrm{p}} B$ と書くことにする。

なお、$\mathcal{UNIQUENUMBER}$ を初期条件として与えると、任意の r.c. に含まれる初期条件を計算できることはすでにわかっている。

3 主要結果

Proposition 3.1 匿名ネットワーク上で頂点数の上限が分かっている時、頂点数の約数個の同型な生成木を作る決定性のアルゴリズムが存在する。

これは [坂本 97] で構成した統合化アルゴリズムを利用し、

- 最初は各頂点が頂点数 1 の根付の生成木であるとする。
- 生成木の根が生成木を代表し、生成木が一つの頂点として機能し、隣接する生成木同士と通信を行う。
- 生成木に同型でなければ等しくないような一意な全順序を与え、隣接する生成木の中で最大なものが、隣接する自分より小さい生成木を統合して行く。

という操作で、全ての生成木が同型なるまで統合を繰り返すことにより得られる。

頂点数の上限が分かっていれば、全ての生成木が同型か否かの終了条件が判断できる。

Theorem 3.2 $\mathcal{UNIQUENUMBER} \leq_{\mathrm{zp}} \mathcal{SIZE}$

Proof. [BZ89] により、全頂点数が N で、各頂点が 1 から N^3 までの値をランダムに選んだ時、全ての頂点で選んだ値が異なる確率は

$$
\begin{aligned}
\left(1-\frac{1}{N^3}\right)^{(N-1)N} &= \left((1-\frac{1}{N^3})^{N^3-1}\right)^{\frac{N^2-N}{N^3-1}} \\
&\geq e^{-\frac{N^2-N}{N^3-1}} \\
&= e^{-\frac{N}{N^2+N+1}} \\
&> e^{-\frac{1}{N}}
\end{aligned}
$$

となる。よって、そのように各頂点で乱数を選び、全ての頂点が頂点数 1 の木の根でとして、統合化アルゴリズムを実行すると、全ての乱数が異なる時は一本の生成木になる。与えられた N と生成木の頂点数を比較し、等しければ停止し、異なればこのプロセスを繰り返すようにすれば、確率 1 で $\mathcal{UNIQUENUMBER}$を出力できる。 □

Corollary 3.3 $\mathcal{UNIQUENUMBER} \leq_{\mathrm{p}} \mathcal{UPPERBOUND}$

Proof. 各頂点に入力される値を U とする。頂点数の上限が分かっているので、ネットワーク全体にメッセージを送ることができる。それを利用して、各々入力された U の中で最小の U を求めておく。(最小の U を U' とする)

許されるエラーの確率が r の時、c を

$$
c > \frac{\log r}{\log(1-e^{-1/U'})}
$$

なる値として、定理 3.2 の、乱数を発生し生成木を作るアルゴリズムを c 回実行する。すると c 回とも生成木が一本にならない確率は $(1-e^{-1/U'})^c$ 以下で、これは c の選び方より r より小さくすることができる。よってエラー確率 r で各頂点に異なる番号を振ることが可能である。 □

Definition 3.4 S を自然数の帰納的可算集合とした時、S-$\mathcal{SIZE}$ は S のある要素 x とネットワークのサイズ N の積 xN を全頂点に与える初期条件とする。

Theorem 3.5 T-$\mathcal{SIZE} \leq_{\mathrm{zp}} S$-$\mathcal{SIZE} \Leftrightarrow (\forall x \in S)$ $(\exists c > 0)$ $cS_x \subseteq T$ *where* $S_x = \{y \mid y|x,\ y \in S\}$

Proof. $\Leftarrow$

頂点数 N のネットワーク上の分散アルゴリズムの入力に xN $(x \in S)$ が与えられた時、 $(xN)^3$ までの乱数を発生し、 proposition 3.1のアルゴリズムにより生成木を作る。すると、生成木内の頂点数として、 $n = N/d$ (d はある整数) が得られる。この時、 S と、与えられた xN と、計算により求めた n に対して $xN = x_i n$ $(x_i \in S)$ となるような x_i が見つかるまでこれを繰り返す。

x_i が見つかった時、全頂点数 N は $\frac{x_i}{x}n = \frac{x_i}{dx}N$ と考えられる。つまり、 $x = x_i/d$。 S_{x_i} の要素 y_1, y_2, ..., $y_k = x_i$ に対して、最初に与えられた xN は、ある j $(1 \leq j \leq k)$ に対して $y_j N$ となる。その時、 $cS_{x_i} = c\{y_1, \ldots, y_k\} \subseteq T$ となる c が仮定より存在し、 S, T が帰納的可算集合なので、 c を effective に求めることが可能である。よって、出力として、 $cx = cy_j \in T$ より、 cxN を出力することで $T\text{-}\mathcal{SIZE}$ を満たす。

$\Rightarrow$

仮定より、 S は、ある $x \in S$ に対し、すべての $c > 0$ に対して $cS_x \not\subseteq T$ が成り立つとする。その条件を満たす $S_x = \{x_1, \ldots, x_{k-1}, x_k = x\}$ に対して、頂点数 $3, 3x/x_{k-1}, \ldots, 3x/x_1$ のリングを考える。各リングに $3x$ という初期値を与えるのは $S\text{-}\mathcal{SIZE}$ の初期条件を満たす。

さて、上記の条件を満たす S, T に対して、ある分散アルゴリズム M によって $T\text{-}\mathcal{SIZE} \leq_{\mathrm{zp}} S\text{-}\mathcal{SIZE}$ が成り立つと仮定する。

$3x$ を与えた頂点数 3 のリングにおいて M を実行した時、仮定より、その出力が $\{3y_1, 3y_2, \ldots\}$ で、 $\{y_1, y_2, \ldots\} \subseteq T$ であるとする。xその時、 y_1 を出力するときに使用した乱数列はそれぞれ有限なので、その乱数列をサイズ $3x/x_{k-1}$, ..., $3x/x_1$ のリングの各頂点に3つおきに与えると、 $3y_1$ を出力する。この乱数列を生成する確率はそれぞれのリング上で有限の確率になる。さて、この出力が $T\text{-}\mathcal{SIZE}$ の初期条件を満たす条件は、 $\frac{3y_1}{3\frac{x}{x_i}} \in T$ となることであるが、 $\frac{y_1}{x}\{x_1, x_2, \ldots, x_k = x\} \subseteq T$ となるためには、 $\frac{y_1}{x}S_x \subseteq T$ となることが必要であり、これは、仮定に矛盾する。 $\square$

ここで、この、 $(\forall x \in S)\ (\exists c > 0)\ \{cy \mid y|x,\ y \in S\} \subseteq T$ を $S \preceq T$ と表すと、この関係 $\preceq$ は反射律、推移律を満たす。また、 $\sim$ を $\preceq$ の作る同値関係、 $A \preceq B$ かつ $A \not\sim B$ を $A \prec B$ と表す。

$\preceq$ の基本的な性質として、

1. $A \subseteq B$ ならば $A \preceq B$
2. $\{1, s\}$, $\{1, t\}$ は $s \neq t$ ならば $\{1, s\} \not\preceq \{1, t\}$ かつ $\{1, t\} \not\preceq \{1, s\}$
3. $\{1\} \prec \{1, 2\} \prec \{1, 2, 3\} \prec \cdots$
4. $S \sim cS$

が成り立つ。

参考文献

[BZ89] J. Bar-Ilan and Dror Zernik. Random leaders and random spanning trees. *Lecture Notes in Computer Science*, Vol. 392, pp. 1–12, 1989. Distributed Algorithms.

[HR90] T. Hagerup and C. Rüb. A Guided Tour of Chernoff Bound. *Information Processing Letters*, Vol. 33, pp. 305–308, 1989/90.

[坂本 97] 坂本直志. 分散アルゴリズムの初期条件の作る構造について. 情報処理学会研究報告, Vol. 97, No. 115, pp. 41–48, 1997. 97–AL–59–6.